

# **Universiti Teknologi MARA**

## **DATA VISUALIZATION OF CIGARETTE COMMENTS FROM SOCIAL MEDIA**

**NURUL NAZIRAH BINTI MOHAMAD NASIR**  
**2015666706**

**Thesis submitted in fulfillment of the requirements for  
Bachelor of Computer Science (Hons.)  
Faculty of Computer and Mathematical Sciences**

**July 2017**

## **SUPERVISOR'S APPROVAL**

### **DATA VISUALIZATION OF CIGARETTE COMMENTS FROM SOCIAL MEDIA**

By

**NURUL NAZIRAH BINTI MOHAMAD NASIR**  
**2015666706**

This thesis was prepared under the direction of thesis supervisor, Encik Ahmad Firdaus bin Ahmad Fadzil. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons).

Approved by:

.....

AHMAD FIRDAUS BIN AHMAD FADZIL  
Project Supervisor

JULY 24, 2017

## **STUDENT'S DECLARATION**

I certify that this report and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

.....

NURUL NAZIRAH BINTI MOHAMAD NASIR

2015666706

JULY 24, 2017

## ABSTRACT

Cigarette is the small roll of paper that is filled with cut tobacco for smoking. Nowadays, many people use the social media to express their feeling in something, share information, give opinion or changes idea each other through the Facebook and Twitter. This study is analyze the data of cigarettes that people comments in social media which are in Facebook, Twitter, and Google search engine. Currently, raw data extracted from the said location contains a lot of info, but mostly vague and difficult to interpret. There is a lot of data on cigarettes that people comments in social media. Then, data comments that in each word usually has same word but has different meaning. This made the task to understand the data difficult as there are no clarity regarding the comments. In order to solve these problem, the visualization will be used to present the data of cigarette. The data from social media will be match with Google correlate. All information that will be then converted into multidimensional visualization which are bubble chart and bubble cloud. To visualize the data, D3 Java library and Python server platform is employed. The preprocessing of the data is executed by using C# language. The result indicated that the visualization able to properly visualize the comments regarding cigarette effectively. Response from users shows positive feedback.

**Keywords:** Cigarette comments, Social Media, Multidimensional, D3, Python and C# Programming Language.

## TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR'S APPROVAL	ii
STUDENT'S DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xii
CHAPTER ONE: INTRODUCTION	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Research Objectives	3
1.4 Research Scope	3
1.5 Research Significances	4
1.6 Summary	4
CHAPTER TWO: LITERATURE REVIEW	5
2.1 Introduction	5
2.2 Cigarettes and social media	5
2.2.1 Cigarette in General	6
2.2.2 Comments on cigarette in social media	6
2.2.3 Mining data from social media	7
2.3 Processing vague and raw data	8
2.3.1 Data cleaning and filtering	8
2.3.2 Data classification	8
2.4 Data Visualization	9
2.4.1 Type of Visualization	9
2.4.2 Comparison between Previous Projects	19
2.5 Summary	22